


SANDIA REPORT

SAND98-1109

Unlimited Release

Printed September 1999



Small Sample Properties of an Adaptive Filter with Application to Low-Volume Statistical Process Control

Stephen V. Crowder

Prepared by

Sandia National Laboratories

Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of
Energy under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.



Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Prices available from (703) 605-6000
Web site: <http://www.ntis.gov/ordering.htm>

Available to the public from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161

NTIS price codes
Printed copy: A04
Microfiche copy: A01



SAND98-1109
Unlimited Release
Printed September 1999

Small Sample Properties of an Adaptive Filter with
Application to Low Volume Statistical Process Control

Stephen V. Crowder
Statistics and Human Factors Division
Sandia National Laboratories
P. O. Box 5800
Albuquerque, NM 87185

Abstract

In many manufacturing environments such as the nuclear weapons complex, emphasis has shifted from the regular production and delivery of large orders to infrequent small orders. However, the challenge to maintain the same high quality and reliability standards while building much smaller lot sizes remains. To meet this challenge, specific areas need more attention, including fast and on-target process start-up, low volume statistical process control, process characterization with small experiments, and estimating reliability given few actual performance tests of the product. In this paper we address the issue of low volume statistical process control. We investigate an adaptive filtering approach to process monitoring with a relatively short time series of autocorrelated data. The emphasis is on estimation and minimization of mean squared error rather than the traditional hypothesis testing and run length analyses associated with process control charting. We develop an adaptive filtering technique that assumes initial process parameters are unknown, and updates the parameters as more data become available. Using simulation techniques, we study the data requirements (the length of a time series of autocorrelated data) necessary to adequately estimate process parameters. We show that far fewer data values are needed than is typically recommended for process control applications. We also demonstrate the techniques with a case study from the nuclear weapons manufacturing complex.

TABLE OF CONTENTS

Introduction.....	5
Process Model.....	6
Estimation With Parameters Known.....	7
Properties of Estimator With Parameters Known.....	8
Estimation With Parameters Not Known.....	11
Properties of Two-Stage Estimator With Parameters Not Known.....	12
Application to SPC and LVSPC.....	13
Case Study Using Battery Cathode Data.....	15
Summary and Conclusions.....	16
Acknowledgements.....	17
Appendix A-Estimation of Mean Vector with Variance Components Known.....	17
Appendix B-Estimation of Variance Components Via Maximum Likelihood.....	18
Appendix C-Extension to Other Models.....	20
References.....	23
Table 1.-Increase in MSE as a Function of Sample Size.....	25
Figure 1.-Simulations of Process Model for $\Theta=.05$, 1.0, and 5.0.....	26
Figure 2.-Weights Applied to Process Data for $\Theta=.05$	27
Figure 3.-Weights Applied to Process Data for $\Theta=1.0$	28
Figure 4.-Weights Applied to Process Data for $\Theta=5.0$	29
Figure 5.-Simulated Distributions for $\Theta=.05$, with Sample Size $n=10, 25, 50, 75, 100$, and 200.....	30
Figure 6.-Simulated Distributions for $\Theta=1.0$, with Sample Size $n=10, 25, 50, 75, 100$, and 200.....	31
Figure 7.-Simulated Distributions for $\Theta=5.0$, with Sample Size $n=10, 25, 50, 75, 100$, and 200.....	32
Figure 8.-Battery Cathode Weights from a Low Volume Production Run.....	33
Figure 9.-Battery Cathode Weights with Confidence Intervals ($n=25$ and $n=50$).....	34
Figure 10.-Battery Cathode Weights with Confidence Intervals ($n=75$ and $n=100$).....	35
Figure 11.-Battery Cathode Weights with Confidence Intervals ($n=143$).....	36
Figure 12.-Estimates of Θ for Sample Sizes $n=3, 4, 5, \dots, 143$	37

Introduction

Advanced manufacturing technology, pressured by the need for greater efficiency and responsiveness to the customer, now relies heavily on strategies such as low volume production runs to more quickly meet the changing demands of the customer and to lower manufacturing costs. In many industries this situation has included a dramatic shift in production philosophy from the delivery of a few big orders to many smaller ones. As a result, there is a need to develop statistical process control (SPC) techniques that are useful in low volume manufacturing environments, where relevant data may be scarce. This is especially true in the nuclear weapons complex, with shrinking budgets and decreasing need for large quantities of new weapons. For the small lot manufacturing that remains, it is particularly important to have modern, appropriate statistical methods in place to support and enhance quality improvement efforts.

In a low volume manufacturing environment, the data necessary to accurately estimate process parameters such as the process mean and standard deviation, and the limits for standard statistical control charts, are usually not available prior to the start of production. The usual recommendation for establishing valid control limits for the \bar{X} chart, for example, is to gather around 25 samples of 4 to 6 observations each as a baseline (see Montgomery (1991)). Quesenberry (1993) recommends a sample of size 300 to establish permanent limits for the individuals chart. Samples this large may not be feasible in a low volume environment, since the number of data points needed to satisfy these recommendations might exceed the total number of parts produced in a short production run. Various techniques that have been proposed to address this problem are summarized in a report by Crowder and Halbleib (1995).

In this paper we investigate an adaptive filtering approach to process monitoring with a relatively short time series of autocorrelated data. We consider the cases in which the model parameters are known and unknown, and include a discussion of the associated estimators in each case. For the case in which the parameters are known, we describe the estimation procedure and discuss how the filtering algorithm uses the data from the time series, as well as other properties of the resulting Bayesian estimator. For the case in which the parameters are not known, more typical in a low volume manufacturing environment, we describe a two-stage estimation procedure combining maximum likelihood estimation in the first stage with Bayesian estimation in the second stage. This adaptive procedure assumes that initial process parameters are not known, and updates the estimates as more data become available.

Much has appeared in the recent literature on process control charting with autocorrelated data. See Vander Wiel (1996), and Wardell, et al (1992, 1994), for example. Most of the attention has been given to monitoring the residuals from a specified time series model to detect mean level shifts, with emphasis on average run lengths. Little attention has been given to the problem of sample size requirements for possibly short time series associated with low volume production. The recommendations above for very large sample sizes

are based on how large the sample size must be for the estimated control limits to perform essentially like the known control limits. The emphasis here, however, is on estimation and minimization of mean squared error (MSE) rather than the traditional hypothesis testing and run length analyses associated with control charting and control limits. With low volume production and small lots, hypothesis testing and run length analyses may not even be meaningful, since the expected run length may well exceed the entire production quantity. Using simulation techniques, we investigate the small sample properties of a two-stage estimation procedure by evaluating mean squared errors and convergence rates of parameter estimates. Based on the results of this investigation we make recommendations regarding sample size requirements for estimating parameters from a possibly short time series of autocorrelated data. We show that far fewer data values are needed than is typically recommended for control charting applications. We also demonstrate the techniques with a case study from the nuclear weapons manufacturing complex, applying the adaptive filtering algorithm to a time series of battery cathode weight data. Finally we discuss extensions to more general time series models.

Process Model

The model that we consider for monitoring a process mean with a possibly small sample is the so-called steady model (see Smith (1979)). The scalar steady model is described by

$$y_t = \mu_t + \varepsilon_t \quad \text{with the } \varepsilon_t \text{'s i.i.d. } N(0, \sigma_\varepsilon^2) \quad (1)$$

$$\mu_t = \mu_{t-1} + v_t \quad \text{with the } v_t \text{'s i.i.d. } N(0, \sigma_v^2)$$

$$t = 1, 2, \dots, n \quad \mu_0 \text{ is the initial mean}$$

where μ_t is the true unknown process mean at time t , and $\{y_t, t = 1, 2, \dots, n\}$ are the observed process outputs. The sequence $\{\varepsilon_t, t = 1, 2, \dots, n\}$ can be thought of as random observation errors while the sequence $\{v_t, t = 1, 2, \dots, n\}$ can be thought of as random shocks to the process.

It is convenient to describe this process in terms of the ratio of the two variance

components, which we denote $\theta = \frac{\sigma_v^2}{\sigma_\varepsilon^2}$. Figure 1 shows simulations of the above process

model for $\theta = .05, 1.0$, and 5.0 , each with $\sigma_\varepsilon = 1.0$. Note that the volatility of the process is a function of the parameter θ . With $\theta = .05$ the process is relatively stable, while with $\theta = 5.0$, the process is very unstable. Note that $\theta = 0$ in model (1) corresponds to an independent and identically distributed (i.i.d) process model, often used in SPC applications. In practice it is expected that θ will typically fall between $.05$ and 1.0 , a range associated with mild to moderate levels of autocorrelation.

[Insert Figure 1 Here]

Note that for this model the sequence μ_t is a random walk, appropriate for situations where the most important characteristic of the process in question is its current mean level, with persistent growth or decline being either absent or unimportant. Harrison and Stevens (1976) discuss this model as a special case of their Dynamic Linear Model. Harrison (1967) uses the model to describe customer demand for a steady selling product, and as a tool in short-term sales forecasting. Meinhold and Singpurwalla (1983) examine this model as a special case in their overview of the Kalman filter. One of the advantages of this representation is the parametric structure of the model. At any given time, probabilistic information regarding each μ_t is available in the form of a posterior distribution given all available data. The model representation (1) also describes how the process mean changes over time as a result of random shocks.

It should also be pointed out that the steady model has the same autocorrelation structure as a class of ARIMA(0, 1, 1) models of Box and Jenkins (1970), models that are widely used in practice. This can be shown by looking at the correlation between successive differences $(y_t - y_{t-1})$ and $(y_{t-1} - y_{t-2})$. Box and Jenkins mention that models of this kind have been found useful in inventory control problems, in representing some kinds of disturbances occurring in industrial processes, and in econometrics. In this paper we first discuss estimation under model (1) with the parameters μ_0 , σ_ϵ^2 , and σ_v^2 known. We then illustrate a two-stage procedure for estimation under model (1) with the parameters μ_0 , σ_ϵ^2 , and σ_v^2 unknown, and study the small sample properties of these estimates, relevant to the problem of low volume statistical process control.

Estimation with Parameters Known

The derivation of the best estimates of the process mean uses a Bayesian approach. We first assume that the initial process mean μ_0 and the variance terms σ_ϵ^2 and σ_v^2 are known. In practice, especially with a low volume or startup process, these parameters will not be known. In a later section we address the same estimation problem with these parameters unknown and study the properties of the resulting estimators.

With $\underline{y} = (y_1, y_2, \dots, y_n)'$ the data vector and $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_n)'$ the mean vector from model (1), we wish to derive $E(\underline{\mu} | \underline{y})$, the posterior mean and $Var(\underline{\mu} | \underline{y})$, the posterior variance of the vector $\underline{\mu}$. The posterior mean will be our estimate of $\underline{\mu}$ and the posterior variance will provide an uncertainty band about that estimate on the resulting control chart. The derivation, given in Appendix A, uses the fact that the conditional distribution of $\underline{y} | \underline{\mu}$ is multivariate normal with mean $\underline{\mu}$ and variance-covariance matrix $I\sigma_\epsilon^2$ while the mean vector $\underline{\mu}$ has a multivariate normal distribution that depends only on μ_0 and

σ_v^2 . The conditional distribution of $\underline{y} | \underline{\mu}$ is used as a likelihood function and the distribution of $\underline{\mu}$ is used as a “prior” to develop the Bayesian estimator of the mean vector $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_n)'$. Using these results (see Appendix A), the posterior distribution of $\underline{\mu} | \underline{y}$ is also multivariate normal with posterior mean

$$\begin{aligned} E(\underline{\mu} | \underline{y}) &= \underline{1}\mu_0 + V\sigma_v^2(V\sigma_v^2 + I\sigma_\varepsilon^2)^{-1}(\underline{y} - \underline{1}\mu_0) \\ &= \underline{1}\mu_0 + V\theta(V\theta + I)^{-1}(\underline{y} - \underline{1}\mu_0) \end{aligned} \quad (2)$$

and posterior variance

$$\begin{aligned} Var(\underline{\mu} | \underline{y}) &= V\sigma_v^2 - V\sigma_v^2(V\sigma_v^2 + I\sigma_\varepsilon^2)^{-1}V\sigma_v^2 \\ &= V\theta(V\theta + I)^{-1}\sigma_\varepsilon^2 \end{aligned} \quad (3)$$

using the substitution $\theta = \frac{\sigma_v^2}{\sigma_\varepsilon^2}$. The matrix V in expressions (2) and (3) is defined by

$V_{ij} = \min(i, j)$ for all $i, j \in \{1, \dots, n\}$. The matrix I is the $n \times n$ identity matrix, and $\underline{1}$ is an $n \times 1$ vector of ones.

Since it is the posterior mean, $E(\underline{\mu} | \underline{y})$ is thus the Bayes estimator (under squared error loss) of the vector of μ_i 's. Note that with μ_0 , σ_ε^2 , and σ_v^2 known, the posterior variance does not depend on the data. Note also that the weights applied to the data vector \underline{y} , contained in the matrix expression $V\theta(V\theta + I)^{-1}$, depend only on θ and the sample size n . In the next section we will investigate these weights and the properties of the estimator $E(\underline{\mu} | \underline{y})$ further.

Properties of Estimator With Parameters Known

The posterior mean $E(\underline{\mu} | \underline{y})$ is the Bayes estimator, under squared error loss, of the vector of μ_i 's (assuming normality of the ε_i 's and the v_i 's) and thus has the optimality properties of a Bayes estimator (see Sarris (1973)). When μ_0 , σ_ε^2 , and σ_v^2 are known, these results hold for *any* sample size. The posterior mean $E(\underline{\mu} | \underline{y})$ is also the Kalman filter estimate of the mean vector $\underline{\mu}$. See Meinhold and Singpurwalla (1983) for a discussion of the Kalman filter as a Bayesian estimate. Duncan and Horn (1972) show that even if the ε_i 's and the v_i 's are not normally distributed, the Kalman filter estimator

will still be the minimum mean square *linear* estimator provided the ε_i 's and the v_i 's are independent with zero means and finite second moment.

Our approach is to use $E(\underline{\mu} | \underline{y})$ to update and improve the estimate of *each* μ_i after *every* new data point is obtained. For a low volume process this feature is especially important because it allows improvement of initial estimates based on samples that are possibly very small.

Figures 2-4 show how the weights applied to the process data vary for $\theta = .05, 1.0$, and 5.0 . These figures show the relative weights for the case in which a time series of 20 observations is available for estimating a mean vector. Again, these weights can be computed *before* any data are collected if θ is known. If θ is not known, the weights will be estimated adaptively.

[Insert Figures 2-4 Here]

Figure 2 shows the weights for estimating μ_1 , μ_{10} , and μ_{20} when $\theta = .05$. When θ is relatively small, the process mean changes very little from one time period to the next (refer to Figure 1). So for a short time series, when estimating μ_1 , the initial mean μ_0 receives most of the weight. The weights applied to the data vector $\underline{y} = (y_1, y_2, \dots, y_{20})'$ are largest for the first observation y_1 and decrease slightly for each successive data value. But all twenty observations are used for estimating μ_1 as well as the other μ_i 's. To estimate μ_{10} the initial mean receives very little weight, as expected, and observation y_{10} receives the greatest amount of weight. The weights then decrease in both directions away from y_{10} . To estimate μ_{20} the initial mean again receives very little weight, with observation y_{20} receiving the greatest amount of weight and each preceding data value receiving less weight. Note that this weighting scheme differs from traditional SPC charting which would use only data before or at time t to estimate μ_t .

It can also be shown (see Crowder (1986)) that the estimate of the *last* element in the mean vector, $E(\mu_n | \underline{y})$, is an exponentially weighted moving average (EWMA) of the initial mean μ_0 and the data $(y_1, y_2, \dots, y_n)'$, with EWMA parameter

$$\lambda = \frac{(-\theta + (\theta^2 + 4\theta)^{1/2})}{2}. \quad (4)$$

With the parameter $\theta = .05$ in expression (4) this corresponds to $\lambda = .20$, $\theta = 1.0$ corresponds to $\lambda = .62$, and $\theta = 5.0$ corresponds to $\lambda = .85$. In practical applications a value for λ between .2 and .4 is often recommended.

Figure 3 shows the weights for estimating μ_1 , μ_{10} , and μ_{20} when $\theta = 1.0$. In this figure the pattern of the weights is the same as discussed above, but the relative weights are

much different. When $\theta=1.0$, the process mean changes more from one time period to the next (refer to Figure 1). So as expected, to estimate μ_i , observation y_i receives more weight when θ is larger. Figure 4 shows the weights for estimating μ_1 , μ_{10} , and μ_{20} when $\theta=5.0$. In this figure the pattern of the weights is again the same as discussed above, but now to estimate μ_i , observation y_i receives almost all the weight. When θ is this large, the process mean can change dramatically from one time period to the next (refer to Figure 1), so the weighting pattern in Figure 4 is expected. So when θ is known, the weights should be investigated before any data are collected. In low volume SPC this is especially important because the weights show what data contain most of the information about each process mean μ_i . When θ is large, for example, only a few observations before and after y_i contain information about μ_i . A decision about the process mean during a short production run could thus be delayed until all relevant data were available, and the best possible estimate could be obtained without unnecessary delay.

When θ is essentially zero ($\sigma_v \ll \sigma_e$), using expressions (2) and (3) above,

$$E(\mu_i | \underline{y}) \equiv \mu_0, \text{ the initial mean, for every } i, \text{ and}$$

$$Var(\mu_i | \underline{y}) \equiv i\sigma_v^2.$$

In this extreme case, the estimator essentially ignores the data since σ_e , the observation error standard deviation, is very large relative to σ_v , the system standard deviation. This means that the process mean is nearly constant and equal to μ_0 , the initial mean, and most of the variation in the data is due to observation error, not movement of the process mean. So the estimate of the process mean should remain nearly constant.

In the other extreme case, when θ is very large ($\sigma_v \gg \sigma_e$), again using expressions (2) and (3) above,

$$E(\mu_i | \underline{y}) \equiv y_i, \text{ the } i \text{ th observation, for every } i, \text{ and}$$

$$Var(\mu_i | \underline{y}) \equiv \sigma_e^2 \text{ for every } i.$$

In this case, the estimator uses only the most current data since σ_e , the observation error standard deviation, is small relative to σ_v , the system standard deviation. So the process mean is changing dramatically, and because σ_e is small most of the information about the current process mean is contained in the current observation. Most of the variation in the data is due to random shocks to the process mean, not observation error. With essentially no observation error, the best estimate of the current process mean should be

the most recent data value. In most cases, however, it is our experience that θ will fall somewhere between .05 and 1.0 and the weights applied to the data will be as they appear in Figure 2 or 3.

Estimation With Parameters Not Known

In this section we present estimates of the process mean vector $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_n)'$ assuming that the initial process mean μ_0 and the variance terms σ_ε^2 and σ_v^2 are unknown. In practice, especially with a low volume or startup process, these parameters will typically not be known. The mean vector will be estimated using a two-stage approach outlined by Sarris (1973). The first stage of the procedure uses maximum likelihood to estimate the unknown parameters μ_0 , σ_ε^2 , and σ_v^2 . The second stage then uses these estimates in the Bayesian estimation procedure described earlier to estimate the unknown mean vector. In the next section we will investigate the important small sample properties of this two-stage procedure and its applicability for low volume statistical process control.

The maximum likelihood equations used to estimate μ_0 , σ_ε^2 , and σ_v^2 are developed in detail in Appendix B. The derivation uses the fact that the vectors $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_n)'$ and $\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ from model (1) both have multivariate normal distributions. It is then the case that $\underline{y} = \underline{\mu} + \underline{\varepsilon}$ also has a multivariate normal distribution that depends on the unknown parameters μ_0 , σ_ε^2 , and σ_v^2 . Its multivariate normal density function is the likelihood function that is used to obtain the maximum likelihood estimates (MLEs) of μ_0 , σ_ε^2 , and σ_v^2 .

Using the maximization techniques outlined in Appendix B, the MLE of μ_0 as a function of θ is

$$\hat{\mu}_0(\theta) = \left(\underline{1}' P^{-1}(\theta) \underline{1} \right)^{-1} \underline{1}' P^{-1}(\theta) \underline{y}, \quad (5)$$

and

$$\hat{\sigma}_\varepsilon^2(\theta) = \frac{1}{n} \left(\underline{y} - \underline{1} \hat{\mu}_0(\theta) \right)' P^{-1}(\theta) \left(\underline{y} - \underline{1} \hat{\mu}_0(\theta) \right) \quad (6)$$

is the maximum likelihood estimate of the observation error variance as a function of θ . The matrix $P(\theta)$ in expressions (5) and (6) is defined by $P(\theta) = (V\theta + I)$. Note that if $\theta=0$ (the i.i.d. case of the steady model) in (5), $P(\theta) = I$, the identity matrix. Then $\hat{\mu}_0(\theta)$ reduces to the usual sample average and $\hat{\sigma}_\epsilon^2(\theta)$ reduces to the usual sample variance with denominator n . The third parameter estimated using the maximum likelihood equations, $\theta = \frac{\sigma_v^2}{\sigma_\epsilon^2}$, has no closed form solution and must be estimated numerically. Once the maximum likelihood estimate of θ is obtained, its value is substituted into expressions (5) and (6) to obtain estimates of μ_0 and σ_ϵ^2 . The maximum likelihood estimates are then substituted into equations (2) and (3) to obtain an estimate of the posterior mean $E(\underline{\mu} | \underline{y})$ and the posterior variance $Var(\underline{\mu} | \underline{y})$. We denote the resulting two-stage estimator of the mean vector $\underline{\mu}$ by $\hat{E}(\underline{\mu} | \underline{y})$, and refer to the resulting procedure of updating after every new observation as an adaptive filter.

Properties of Two-Stage Estimator With Parameters Not Known

Since μ_0 , σ_ϵ^2 , and θ are not known exactly, the error of $E(\underline{\mu} | \underline{y})$ in estimating $\underline{\mu}$ will be increased in the two-stage procedure. Sarris (1973) discusses this procedure in a more general context and Cooley (1971) shows that the estimators of quantities like μ_0 , σ_ϵ^2 , and θ are consistent and efficient, both *large* sample properties. So for large samples, the optimality properties of the Bayesian estimator $E(\underline{\mu} | \underline{y})$ will still hold true. However, the results for small samples are not as well known, and will be investigated further here.

The distributions of the maximum likelihood estimators were studied in detail and the increase in mean squared error (MSE) associated with the two-stage procedure was evaluated, as a function of sample size, using simulation techniques. These results can be used as a guide for deciding what minimum sample sizes are required for statistical process control with autocorrelated data using this procedure. The distribution of $\hat{\theta}$, the maximum likelihood estimator of θ was studied by simulating process model (1) 1000 times for different sample sizes and different values of θ . The graphical results are presented in the form of boxplots in Figures 5-7.

[Insert Figures 5-7 Here]

In Figure 5 with $\theta=.05$ the search routine was confined to the interval $[0, .5]$, in Figure 6 with $\theta=1.0$ the search routine was confined to the interval $[0, 10]$, and in Figure 7 with $\theta=5.0$ the search routine was confined to the interval $[0, 25]$. For samples of size 10 the maximum likelihood function often achieved its maximum on a boundary of the search interval, producing unusual looking boxplots. Clearly, samples of size 10 are not adequate for estimating θ . However, it appears that with samples of size 50 or greater

the distribution of $\hat{\theta}$ is well behaved, with the average value of the simulated distribution close to the actual value of θ . For $\theta = .05$ and $\theta = 5.0$, with sample size 25, the distribution of $\hat{\theta}$ is not well behaved, but for $\theta = 1.0$ and sample size 25 the distribution of $\hat{\theta}$ is reasonably well-behaved.

A second metric to use in the small sample study is the MSE associated with estimating the mean vector $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_n)'$. The quantity $E\{\sum_{i=1}^n (\mu_i - E(\mu_i | \underline{y}))^2\}$, the MSE

when the parameters μ_0 , σ_ϵ^2 , and θ are known, was compared to the quantity

$E\{\sum_{i=1}^n (\mu_i - \hat{E}(\mu_i | \underline{y}))^2\}$, the MSE when the parameters μ_0 , σ_ϵ^2 , and θ are not known and

with the maximum likelihood estimates substituted into the Bayesian estimate (2). Table 1 below shows the increase in MSE as a function of sample size for various values of θ . Each increase in MSE that appears in the table was computed by averaging the results from 1000 simulation trials.

[Insert Table 1 Here]

Note that for sample sizes $n=5, 10$, and 15 , the increase in MSE is at or above 50% for all values of θ , suggesting that these small sample sizes will not provide estimates nearly as accurate as when the parameters μ_0 , σ_ϵ^2 , and θ are known. In the other extreme of the table, samples of size 100 or greater are clearly adequate for estimating μ_0 , σ_ϵ^2 , θ , and the mean vector, with increases in the MSE no greater than 10% for most θ . It is for samples of size 25, 50, and 75 that the conclusions are less clear. For samples of size 25, the increases in MSE are mostly over 30%, while for samples of size 50, the increases in MSE are 12-25% with the minimum occurring around $\theta = 1.0$. And for samples of size 75, the increases in MSE are 8-19% with the minimum occurring between $\theta = .25$ and $\theta = 2.0$. So for samples as small as 50 the procedure works reasonably well for most θ , and it appears that samples of size 25 are required as a minimum, especially if θ is small (close to an i.i.d. process).

In the next section we present the algorithms for statistical process control using the above procedures and model (1).

Application to SPC and LVSPC

In this section we outline the estimation procedures with their application to SPC for both the cases when the parameters μ_0 , σ_ϵ^2 and σ_v^2 are known and when they are not known. Process control decisions are based on probabilistic information regarding the process mean, given in the form of a posterior distribution for the mean vector $\underline{\mu}$. The difference between the approach here and standard SPC is that the emphasis is on estimation rather

than on hypothesis testing. The steady model (1) assumes small frequent changes in the process mean, so the question of whether a major “shift” in the mean has occurred is not relevant. There is no test of hypothesis that the process mean has shifted. The basic strategy is to obtain the best possible estimate of the process mean $\underline{\mu}$, given all available data, in terms of a posterior mean and variance. Then any decision to adjust the process is based on those estimates, not on fixed control limits. Other considerations such as process margin relative to specifications and costs of nonconformance and adjustments would also be part of any decision to make a process change.

The algorithms below are presented in a way that makes them readily applicable to statistical process control. As a result of the analysis of the previous section, we recommend a *minimum of 25 observations* before beginning to use the procedure as a tool for making process adjustments when μ_0 , σ_ϵ^2 and σ_v^2 are not known. The plotting procedure can be initiated, however, with as few as 3 observations when μ_0 , σ_ϵ^2 and σ_v^2 are known. In each case the procedure is updated after *every* new data point.

Case 1. Parameters μ_0 , σ_ϵ^2 and σ_v^2 (or θ) known.

Given the data vector $\underline{y} = (y_1, y_2, \dots, y_n)'$, (starting with $n \geq 3$)

1. Compute: $E(\underline{\mu} | \underline{y})$ and $Var(\underline{\mu} | \underline{y})$ from (2) and (3). These are the posterior mean and variance, respectively, of the vector $\underline{\mu}$.
2. Construct a time-ordered plot of the intervals: $E(\mu_i | \underline{y}) \pm k \cdot Var(\mu_i | \underline{y})^{1/2}$, for $i = 1, 2, \dots, n$, where y_n is the most recent data point. The constant k is chosen from a table of normal probabilities, i. e., $k = 1.96$ provides a roughly 95% confidence interval for μ_i , using the posterior distribution of $\mu_i | \underline{y}$.
3. Compare these intervals to a nominal or target value and make a decision based on this estimated process performance as well as any other cost considerations.
4. Collect the next data point y_{n+1} and return to Step 1 to update the estimates.

The more likely case in a low volume manufacturing environment is that the parameters μ_0 , σ_ϵ^2 and σ_v^2 will not be known.

Case 2. Parameters μ_0 , σ_ϵ^2 and σ_v^2 (or θ) not known.

Given the data vector $\underline{y} = (y_1, y_2, \dots, y_n)'$, (starting with $n \geq 25$)

1. Compute the maximum likelihood estimates: $\hat{\theta}$, $\hat{\mu}_0$, and $\hat{\sigma}_\epsilon^2$.
2. Compute: $\hat{E}(\underline{\mu} | \underline{y})$ and $\hat{Var}(\underline{\mu} | \underline{y})$, the estimates $E(\underline{\mu} | \underline{y})$ and $Var(\underline{\mu} | \underline{y})$ with the maximum likelihood estimates substituted for the unknown parameters.

3. Construct a time-ordered plot of the intervals: $\hat{E}(\mu_i | \underline{y}) \pm k \cdot \hat{Var}(\mu_i | \underline{y})^{1/2}$. Simulation studies were performed to help choose the appropriate value of k for the case of unknown process parameters. For small sample sizes (say n less than 50) a larger k value is needed than for large sample sizes (say n greater than 50). We recommend $k=4$ for small sample sizes and $k=3$ for large sample sizes.
4. Compare these intervals to a nominal or target value and make a decision based on this estimated process performance as well as any other cost considerations.
5. Collect the next data point y_{n+1} and return to Step 1 to update the estimates.

In the next section we illustrate the above algorithm with a case study using battery cathode data.

Case Study Using Battery Cathode Data

Figure 8 shows 143 consecutive measurements of cathode weight from a low volume production run of batteries. The cathode is one component of a high-reliability lithium battery produced for weapons applications. The weight of the cathode must be tightly monitored and controlled because of strict requirements on cathode density. Since this process is typically a low volume production process, we use this data to illustrate how the low volume SPC algorithm works in practice. The steady model (1) was fit to the complete data set and shown to be a reasonable model for this time series.

We start by illustrating the approach assuming that only the first 25 observations are available. The algorithm can be started with just a handful of observations if the variance components σ_e^2 and σ_v^2 and the initial mean μ_0 are known. But if these parameters must be estimated, a minimum of 25 observations is recommended.

Figures 9-11 show the results of the analysis using $n=25, 50, 75, 100$, and 143 observations. In practice the estimate of the parameter θ and the mean vector $\underline{\mu}$ would be updated after *every* new measurement y_n for $n=25, 26, \dots, 143$, and the procedure outlined above would be followed to construct each graph. The estimates of the means, along with error bars, would be computed and plotted after *each* new observation and then compared to a target or acceptable region for the process mean.

[Insert Figures 9-11 Here]

Figures 9-11 show the battery cathode weights plotted along with the corresponding confidence intervals for the μ_i 's overlaid, for $n=25, 50, 75, 100$, and 143. Note that the confidence intervals do not always cover the data. This is because the filtering algorithm produces an estimate of the process mean that “smooths” the data. So if the estimate of the posterior variance is not large, the confidence interval need not cover the data value. In practice the interval estimate for the process mean, rather than the data value, would be used to make decisions regarding process adjustments. Figure 11 shows the results of the

algorithm applied to the entire data set. This graph can be used to retrospectively look at process performance during the short production run. Figure 12 shows how the estimate of θ changes sequentially as the sample size increases, for $n=3, 4, 5, \dots, 143$.

[Insert Figure 12 Here]

For each sample size, the likelihood function for θ was maximized over the interval $[0, 5]$, an adequate search interval for this process. With all the data included, the likelihood function is maximized at $\hat{\theta}=.40$. Notice how the estimate of θ varies about this value before converging to .40 when the sample size is sufficiently large. During periods of process instability early in the time series (for example, around observation 25 and observations 45-55), the estimate of θ fluctuates greatly. The variation in the data at these observations is not explainable by the model and is an indication that a “special cause” may be present in the process. After observation 60, the estimate of θ is relatively stable.

Summary and Conclusions

In this paper we have investigated an adaptive filtering approach to process monitoring with a possibly short time series of autocorrelated data. The emphasis in this approach is on estimation and minimization of mean squared error rather than the traditional hypothesis testing and run length analyses associated with process control charting with control limits. The steady model (1) assumes small frequent changes in the process mean, so the question of whether a major “shift” in the mean has occurred is not considered directly. There is no test of hypothesis that the process mean has shifted. The basic strategy is to obtain the best possible estimate of the process mean $\underline{\mu}$, given all available data, in terms of a posterior mean and variance. Then any decision to adjust the process is based on those estimates, not on control limits.

For the case in which initial process parameters are not known, typical in low volume manufacturing, we described a two-stage estimation procedure that combines maximum likelihood estimation in the first stage with Bayesian estimation in the second stage. The resulting adaptive filter updates the process parameters as each new data value becomes available. Using simulation techniques, we studied the data requirements necessary to adequately estimate process parameters. We have shown that far fewer data values are needed than is typically recommended for process control applications. Depending on the value of θ , as few as 25-50 observations may be adequate, based on MSE comparisons. We also demonstrated the techniques with a case study from the nuclear weapons manufacturing complex, applying the adaptive filtering algorithm to a time series of battery cathode weight data.

Finally, the development given here has assumed the steady model (1), which corresponds to an ARIMA(0, 1, 1) model, using the notation of Box and Jenkins (1970). The methods discussed here can in fact be extended to more general time series models, using the correspondence between state-space models such as the steady model and

ARIMA models (see Appendix C). The advantage of the state-space representation is that probabilistic information about the process mean is readily available in the form of a posterior distribution that can be used to construct confidence bands for the process mean.

Computer programs have been written in FORTRAN, C, and Matlab™ code to implement the two-stage estimation procedure described earlier. These are available upon request from the author.

Acknowledgements

The authors wish to thank the editor and the referees for their comments and suggestions which helped to improve the presentation of this paper.

Appendix A. Estimation of Mean Vector with Variance Components Known

In this appendix we derive the best estimates of the process mean using a Bayesian approach, assuming that the initial process mean μ_0 and the variance terms σ_ϵ^2 and σ_v^2 are known. With $\underline{y} = (y_1, y_2, \dots, y_n)'$, the data vector, and $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_n)'$, the mean vector, we have from the steady model that $\underline{y}|\underline{\mu}, \sigma_\epsilon^2 \sim MVN(\underline{\mu}, I\sigma_\epsilon^2)$, where I is the $n \times n$ identity matrix. This distribution plays the role of likelihood function for $\underline{y} = (y_1, y_2, \dots, y_n)'$.

Now if we express the μ_i 's in terms of μ_0 and the v_i 's, we have

$$\begin{aligned}\mu_1 &= \mu_0 + v_1 \\ \mu_2 &= \mu_0 + v_1 + v_2 \\ \mu_3 &= \mu_0 + v_1 + v_2 + v_3 \\ &\vdots \\ \mu_n &= \mu_0 + v_1 + v_2 + v_3 + \dots + v_n,\end{aligned}$$

so that $\underline{\mu}|\mu_0, \sigma_v^2 \sim MVN(\underline{1}\mu_0, V\sigma_v^2)$, where $\underline{1}$ is an $n \times 1$ vector of ones, and

$$V_{ij} = \min(i, j) \text{ for all } i, j \in \{1, \dots, n\}.$$

This distribution plays the role of a “prior” distribution for $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_n)'$.

The distribution of $\underline{y}|\underline{\mu}, \sigma_\varepsilon^2 \sim MVN(\underline{\mu}, I\sigma_\varepsilon^2)$ is then treated as a likelihood function and the distribution of $\underline{\mu}|\underline{\mu}_0, \sigma_v^2 \sim MVN(\underline{1}\underline{\mu}_0, V\sigma_v^2)$ is treated as a prior distribution in the development of a Bayesian estimator of the vector $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_n)'$. Combining these distributional results, the posterior distribution of $\underline{\mu}|\underline{y}$ is also multivariate normal (see the development in Meinhold and Singpurwalla (1983)) with posterior mean

$$\begin{aligned} E(\underline{\mu}|\underline{y}) &= \underline{1}\underline{\mu}_0 + V\sigma_v^2(V\sigma_v^2 + I\sigma_\varepsilon^2)^{-1}(\underline{y} - \underline{1}\underline{\mu}_0) \\ &= \underline{1}\underline{\mu}_0 + V\theta(V\theta + I)^{-1}(\underline{y} - \underline{1}\underline{\mu}_0) \end{aligned}$$

and posterior variance

$$\begin{aligned} Var(\underline{\mu}|\underline{y}) &= V\sigma_v^2 - V\sigma_v^2(V\sigma_v^2 + I\sigma_\varepsilon^2)^{-1}V\sigma_v^2 \\ &= V\theta(V\theta + I)^{-1}\sigma_\varepsilon^2 \end{aligned}$$

using the substitution $\theta = \frac{\sigma_v^2}{\sigma_\varepsilon^2}$.

Appendix B. Estimation of Variance Components via Maximum Likelihood

To develop the maximum likelihood equations to estimate $\underline{\mu}_0$, σ_ε^2 , and $\theta = \frac{\sigma_v^2}{\sigma_\varepsilon^2}$, we

begin by noting that the vectors $\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_n)'$ and $\underline{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ from the steady model both have multivariate normal distributions. Specifically, $\underline{\mu}|\underline{\mu}_0, \sigma_v^2 \sim MVN(\underline{1}\underline{\mu}_0, V\sigma_v^2)$ and $\underline{\varepsilon}|\sigma_\varepsilon^2 \sim MVN(\underline{0}, I\sigma_\varepsilon^2)$ where $\underline{1}$ is an $n \times 1$ vector of ones, $\underline{0}$ is an $n \times 1$ vector of zeroes, and $V_{ij} = \min(i, j)$ for all $i, j \in \{1, \dots, n\}$. Since these two vectors are independent, it follows that $\underline{y} = \underline{\mu} + \underline{\varepsilon}$ has a multivariate normal distribution as well. That is,

$\underline{y}|\underline{\mu}_0, \sigma_\varepsilon^2, \sigma_v^2 \sim MVN(\underline{1}\underline{\mu}_0, V\sigma_v^2 + I\sigma_\varepsilon^2)$. To express the variance in a more natural way, we will again reparametrize by defining $\theta = \frac{\sigma_v^2}{\sigma_\varepsilon^2}$. We may then characterize this

distribution as $\underline{y}|\underline{\mu}_0, \sigma_\varepsilon^2, \theta \sim MVN(\underline{1}\underline{\mu}_0, P(\theta)\sigma_\varepsilon^2)$ where $P(\theta) = V\theta + I$. The likelihood function obtained from this is

$$f(\underline{y}|\mu_0, \sigma_\epsilon^2, \theta) = \frac{1}{(2\pi)^{\frac{n}{2}} |P(\theta) \sigma_\epsilon^2|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} (\underline{y} - \underline{1}\mu_0)' P^{-1}(\theta) (\underline{y} - \underline{1}\mu_0) \right\}.$$

It is easy to see that the logarithm of this function can be expressed as

$$\begin{aligned} \ell(\mu_0, \sigma_\epsilon^2, \theta) &= \ln f(\underline{y}|\mu_0, \sigma_\epsilon^2, \theta) = -\frac{n}{2} \cdot \ln 2\pi - \frac{1}{2} \cdot \ln |P(\theta)| - \frac{n}{2} \cdot \ln \sigma_\epsilon^2 \\ &\quad - \frac{1}{2\sigma_\epsilon^2} (\underline{y} - \underline{1}\mu_0)' P^{-1}(\theta) (\underline{y} - \underline{1}\mu_0). \end{aligned} \quad (\text{B1})$$

This is the usual log-likelihood function we seek to maximize over each parameter. We first obtain an estimate for μ_0 by expanding the quadratic form in the last term as follows:

$$\begin{aligned} \ell(\mu_0, \sigma_\epsilon^2, \theta) &= -\frac{1}{2\sigma_\epsilon^2} (\underline{y}' - \underline{1}'\mu_0) P^{-1}(\theta) (\underline{y} - \underline{1}\mu_0) + \text{terms unrelated to } \mu_0 \\ &= -\frac{1}{2\sigma_\epsilon^2} (\underline{y}' P^{-1}(\theta) - \underline{1}' P^{-1}(\theta) \mu_0) (\underline{y} - \underline{1}\mu_0) + \dots \\ &= -\frac{1}{2\sigma_\epsilon^2} \left\{ \underline{y}' P^{-1}(\theta) \underline{y} - (\underline{1}' P^{-1}(\theta) \underline{y}) \mu_0 - (\underline{y}' P^{-1}(\theta) \underline{1}) \mu_0 + (\underline{1}' P^{-1}(\theta) \underline{1}) \mu_0^2 \right\} + \dots \\ &= -\frac{1}{2\sigma_\epsilon^2} \left\{ \underline{y}' P^{-1}(\theta) \underline{y} - (\underline{1}' P^{-1}(\theta) \underline{y}) \cdot 2\mu_0 + (\underline{1}' P^{-1}(\theta) \underline{1}) \mu_0^2 \right\} + \dots \end{aligned}$$

where the final equality follows from the fact that $P^{-1}(\theta)$ is symmetric and that the transpose of a scalar is just the scalar itself. Differentiating with respect to μ_0 and equating to zero, we have

$$\frac{\partial \ell}{\partial \mu_0} = \frac{1}{\sigma_\epsilon^2} \left\{ \underline{1}' P^{-1}(\theta) \underline{y} - (\underline{1}' P^{-1}(\theta) \underline{1}) \mu_0 \right\} = 0.$$

Solving this for μ_0 gives us

$$\hat{\mu}_0(\theta) = \left(\underline{1}' P^{-1}(\theta) \underline{1} \right)^{-1} \underline{1}' P^{-1}(\theta) \underline{y}. \quad (\text{B2})$$

Similarly, we differentiate (B1) with respect to σ_ε^2 in order to find its maximum likelihood estimate:

$$\frac{\partial \ell}{\partial \sigma_\varepsilon^2} = -\frac{n}{2\sigma_\varepsilon^2} + \frac{1}{2(\sigma_\varepsilon^2)^2} \left(\underline{y} - \underline{1}\mu_0 \right)' P^{-1}(\theta) \left(\underline{y} - \underline{1}\mu_0 \right) = 0$$

which, upon solving for σ_ε^2 , gives us

$$\hat{\sigma}_\varepsilon^2(\theta) = \frac{1}{n} \left(\underline{y} - \underline{1}\mu_0(\theta) \right)' P^{-1}(\theta) \left(\underline{y} - \underline{1}\mu_0(\theta) \right). \quad (\text{B3})$$

We can then substitute (B2) and (B3) into (B1) in order to obtain a numerical estimate of θ . We propose constraining the search to a small interval (e.g. [0, 25]) and performing a local maximization of the log-likelihood function, or equivalently finding

$$\min_{\theta} \left[n \cdot \ln \left\{ \left(\underline{y} - \underline{1}\mu_0(\theta) \right)' P^{-1}(\theta) \left(\underline{y} - \underline{1}\mu_0(\theta) \right) \right\} + \ln |P(\theta)| \right]$$

over this domain. A simple golden section search seems appropriate for this task, since the assumption of unimodality over a small interval is not unreasonable. Numerical investigation showed this assumption to be valid for all but the smallest sample sizes.

Appendix C. Extension to Other Models

The development given here has assumed the steady model, which corresponds to an ARIMA(0, 1, 1) model in that the two models have the same autocorrelation structure. In this appendix we show how the results can be extended to more complicated ARIMA models.

MacGregor (1973) shows the correspondence between the ARIMA model representation and the usual control theory state-space representation (see Astrom (1970)) such as used in the steady model. Using the backward shift B-operator notation (defining $B a_t = a_{t-1}$), the general ARIMA model of order (p, d, q) may be written as

$$y_t = \frac{(1 - \theta_1 B - \dots - \theta_q B^q)}{(1 - B)^d (1 - \phi_1 B - \dots - \phi_p B^p)} \varepsilon_t$$

$$= \frac{\theta_q(B)}{\nabla^d \phi_p(B)} \varepsilon_t \quad (C1)$$

where $\{\varepsilon_t; t = 1, 2, \dots\}$ is a white noise sequence (or more generally a sequence of random shocks) with mean zero and variance σ_ε^2 . The numerator consists of a single term, $\theta_q(B)$, a polynomial of moving average terms. The denominator term $\nabla^d \phi_p(B)$ consists of two terms, a backward difference term ∇^d and $\phi_p(B)$, a polynomial of autoregressive terms. The ϕ 's, θ 's and σ_ε^2 are parameters that must be estimated from the data. For most time series encountered in practice the order of p, d, and q rarely exceeds two.

Alternatively, the same model can be given the state-space representation:

$$y_t = H\theta_t + \varepsilon_t$$

$$\theta_t = J\theta_{t-1} + \Gamma\varepsilon_{t-1} \quad (C2)$$

where y_t is a vector of output variables, θ_t is a vector of state variables, $\{\varepsilon_t, t = 1, 2, \dots, n\}$ is a vector-valued white noise sequence with variance-covariance matrix given by $\text{Var}(\varepsilon_t) = V_\varepsilon$, and H , J , and Γ are matrices of parameters. The corresponding ARIMA model can easily be obtained from this representation as (substituting $\theta_t = (I - JB)^{-1} \Gamma B \varepsilon_t$):

$$y_t = [H(I - JB)^{-1} \Gamma B + I] \varepsilon_t. \quad (C3)$$

However, the inverse problem of determining a state-space representation from an ARIMA model is less straightforward because the solution is not unique. MacGregor

(1973) demonstrates explicitly how convenient state-space representations of ARIMA models with a minimum number of parameters can be derived. We will give a simple example below.

Consider the ARIMA(1, 0, 1) model (equivalently ARMA(1,1)), which can be expressed as

$$y_t - \varphi_1 y_{t-1} = \varepsilon_t - \theta_1 \varepsilon_{t-1},$$

or equivalently using the backward difference operator,

$$y_t = \frac{(1 - \theta_1 B)}{(1 - \varphi_1 B)} \varepsilon_t. \quad (C4)$$

Using the technique from MacGregor (1973), an equivalent state-space representation is

$$y_t = x_t + \varepsilon_t \quad (C5)$$

$$x_t = \varphi_1 x_{t-1} + (\varphi_1 - \theta_1) \varepsilon_{t-1},$$

where x_t plays the role of the state variable, the unknown process mean at time t . To see that model forms (C4) and (C5) are equivalent, solve for x_t in expression (C5):

$$(1 - \varphi_1 B)x_t = (\varphi_1 - \theta_1)B\varepsilon_t, \quad \text{or}$$

$$x_t = \frac{(\varphi_1 - \theta_1)B}{(1 - \varphi_1 B)} \varepsilon_t.$$

Substituting into the equation for y_t ,

$$y_t = \frac{(\varphi_1 - \theta_1)B}{(1 - \varphi_1 B)} \varepsilon_t + \varepsilon_t = \frac{(1 - \theta_1 B)}{(1 - \varphi_1 B)} \varepsilon_t,$$

which matches the expression in equation (C4). Once the state space representation of an ARIMA model has been derived, the Bayesian estimation technique presented in this paper can be applied to the state variable x_t . For example, in model (C5) above the parameters θ_1 , φ_1 , σ_ε^2 , and x_0 would all be estimated after *each* new data point became available and the filtering algorithm would be applied in the same way as with the steady model. The advantage of this representation is that probabilistic information about the process mean x_t is readily available in the form of a posterior distribution. So the methodology described above for the steady model can be extended in a natural way to the general ARIMA(p, d, q) model.

Reference List

- Astrom, K. J. (1970). An Introduction to Stochastic Control Theory. Academic Press, London.
- Box, G. E. P. and G. M. Jenkins (1970). Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.
- Cooley, T. F. (1971). Estimation in the Presence of Sequential Parameter Variation. Unpublished Ph.D. Dissertation, University of Pennsylvania, Philadelphia, PA.
- Crowder, S. V. (1986). Kalman Filtering and Statistical Process Control. Unpublished Ph.D. Dissertation, Iowa State University, Ames, IA.
- Crowder, S. V. and L. L. Halbleib (1995). Statistical Process Control for Low Volume Manufacturing: A Review. Sandia National Laboratories Internal Document, Albuquerque, NM.
- Duncan, A. J. and S. D. Horn (1972). Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis. *Journal of the American Statistical Association*, v. 67, 815-821.
- Harrison, P. J. (1967). Exponential Smoothing and Short-term Sales Forecasting. *Management Science*, v. 13, 821-842.
- Harrison, P. J. and C. F. Stevens (1971). A Bayesian Approach to Short-term Forecasting. *Operational Research Quarterly*, v. 22, 341-362.
- MacGregor, J. F. (1973). Optimal Discrete Stochastic Control Theory for Process Application. *Canadian Journal of Chemical Engineering*, v. 51, 468-478.
- Meinhold, R. J. and N. D. Singpurwalla (1983). Understanding the Kalman Filter. *American Statistician*, v. 37, 123-127.
- Montgomery, D. C. (1991). Introduction to Statistical Quality Control. John Wiley & Sons, NY.
- Quesenberry, C. P. (1993). The Effect of Sample Size on Estimated Limits for \bar{X} and X Control Charts. *Journal of Quality Technology*, v. 25, 237-247.
- Sarris, A. H. (1973). A Bayesian Approach to Estimation of Time-Varying Regression Coefficients. *Annals of Economic and Social Measurement*, v. 2, 501-523.
- Vander Wiel, S. A. (1996). Monitoring Processes That Wander. *Technometrics*, v. 38, 139-151.

Wardell, D. G., H. Moskowitz, and R. D. Plante (1992). Control Charts in the Presence of Data Correlation. *Management Science*, v. 38, 1084-1105.

Wardell, D. G., H. Moskowitz, and R. D. Plante (1994). Run Length Distributions of Special-Cause Control Charts for Correlated Processes, *Technometrics*, v. 36, 3-17.

Keywords: Autocorrelated Data, Adaptive Filtering

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Table 1. Increase in MSE as a Function of Sample Size for Various Values of θ

$n \backslash \theta$.05	.10	.15	.25	.50	1.0	2.0	3.0	4.0	5.0
5	241%	137%	103%	76%	66%	63%	83%	100%	114%	140%
10	108%	82%	69%	66%	57%	58%	65%	79%	93%	101%
15	66%	61%	56%	52%	48%	47%	52%	58%	74%	80%
25	48%	43%	41%	36%	32%	33%	28%	36%	35%	36%
50	28%	25%	20%	17%	15%	12%	15%	14%	15%	18%
75	19%	14%	12%	10%	8%	8%	10%	9%	10%	12%
100	15%	10%	9%	6%	7%	6%	7%	8%	7%	8%
200	6%	4%	3%	3%	3%	3%	4%	4%	5%	5%

Figure 1. Simulations of Process Model (1) for $\theta = .05, 1.0$, and 5.0 , each with $\sigma_\epsilon = 1.0$.

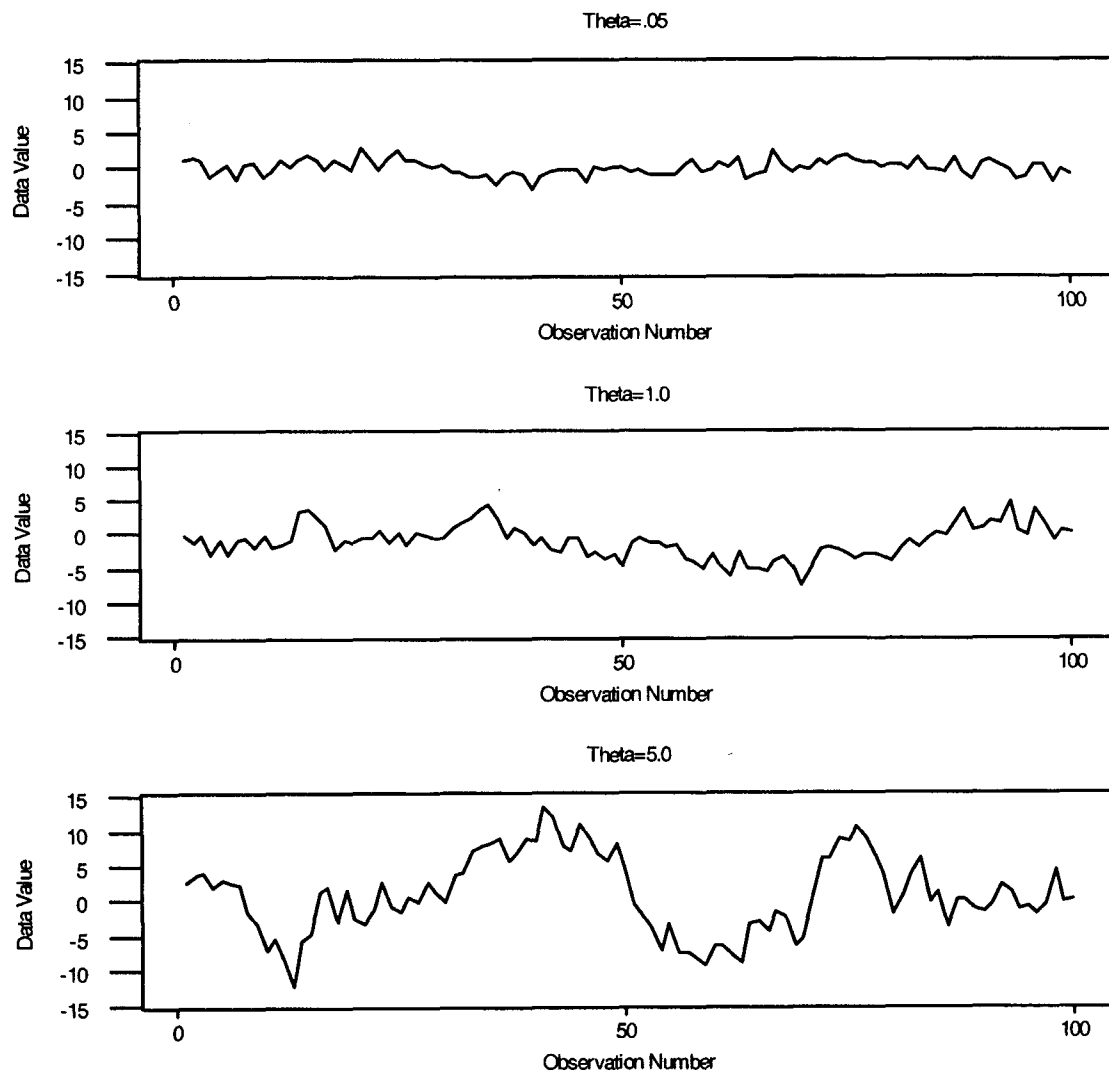


Figure 2. Weights Applied to Process Data for $\theta = .05$.

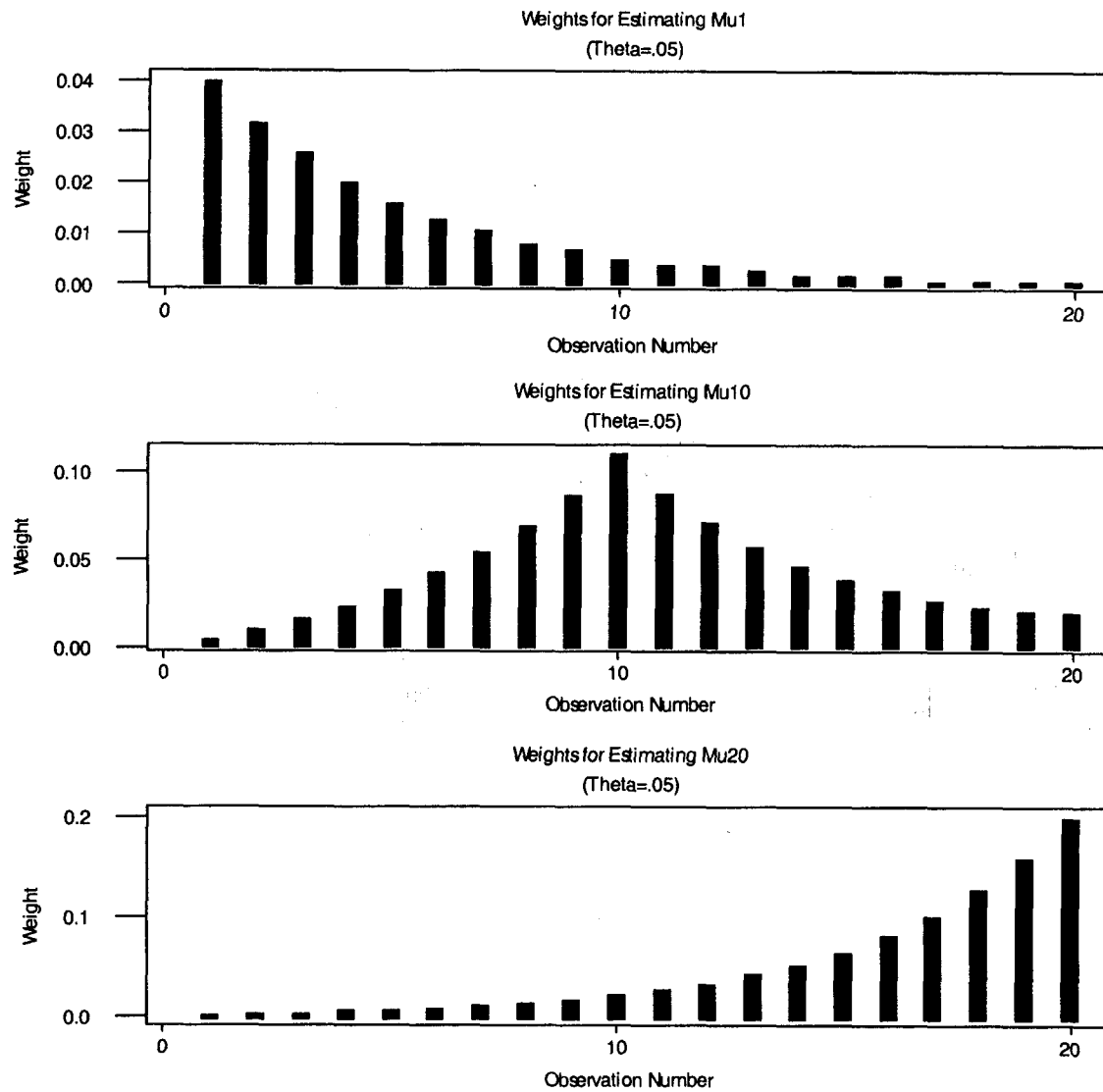


Figure 3. Weights Applied to Process Data for $\theta = 1.0$.

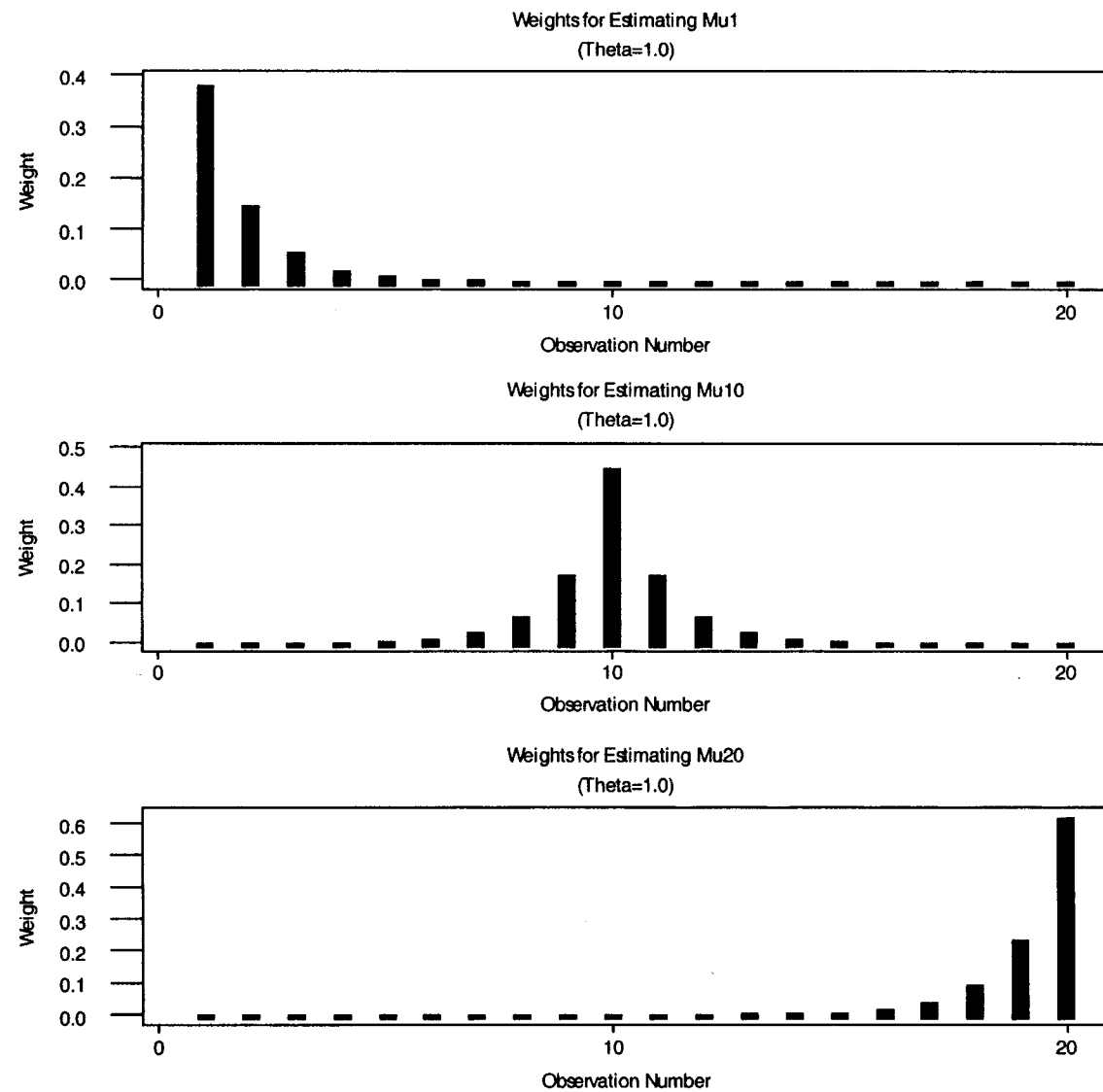


Figure 4. Weights Applied to Process Data for $\theta = 5.0$.

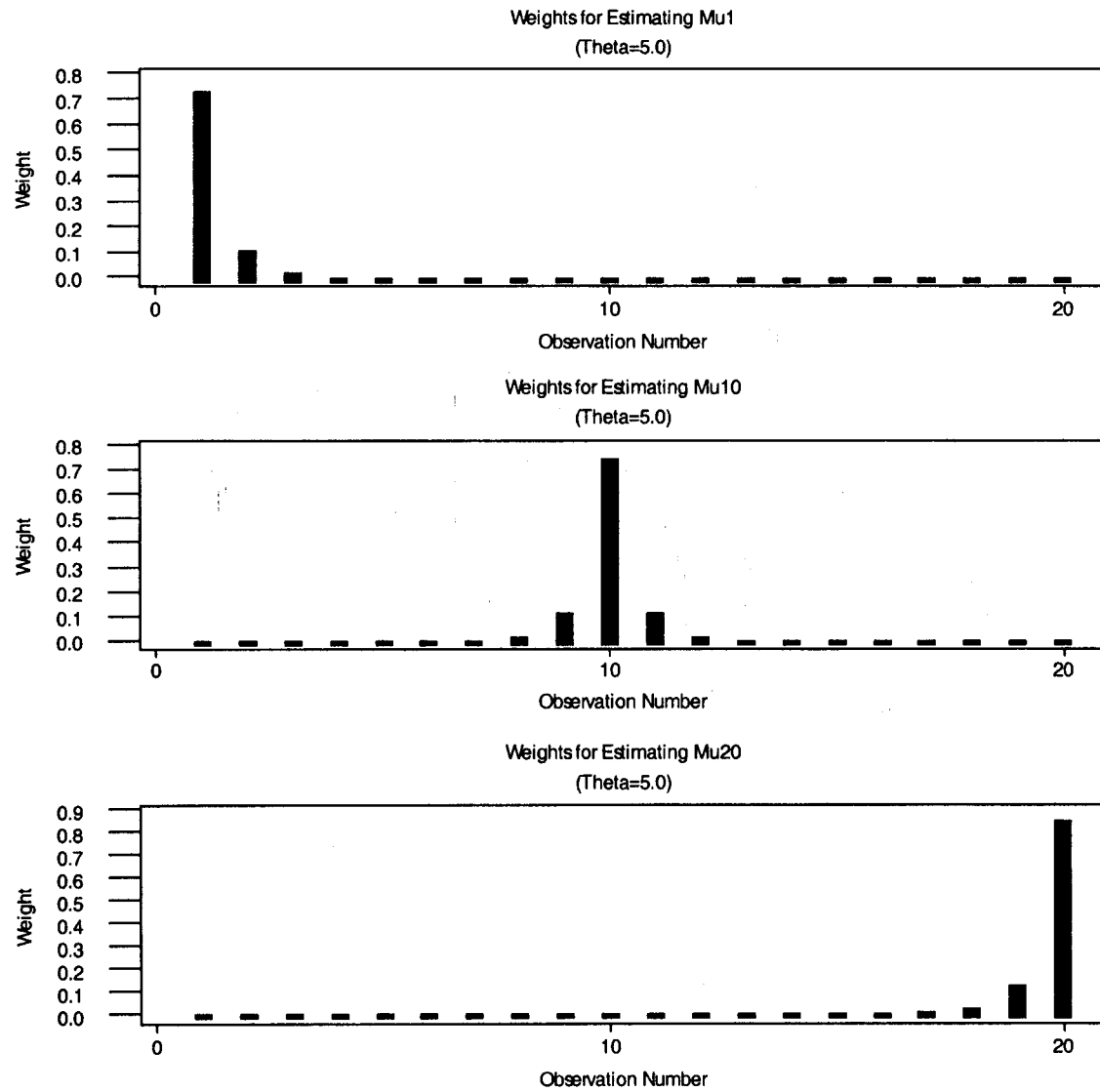


Figure 5. Simulated Distributions of $\hat{\theta}$ for $\theta = .05$, with Sample Sizes $n = 10, 25, 50, 75, 100$, and 200 .

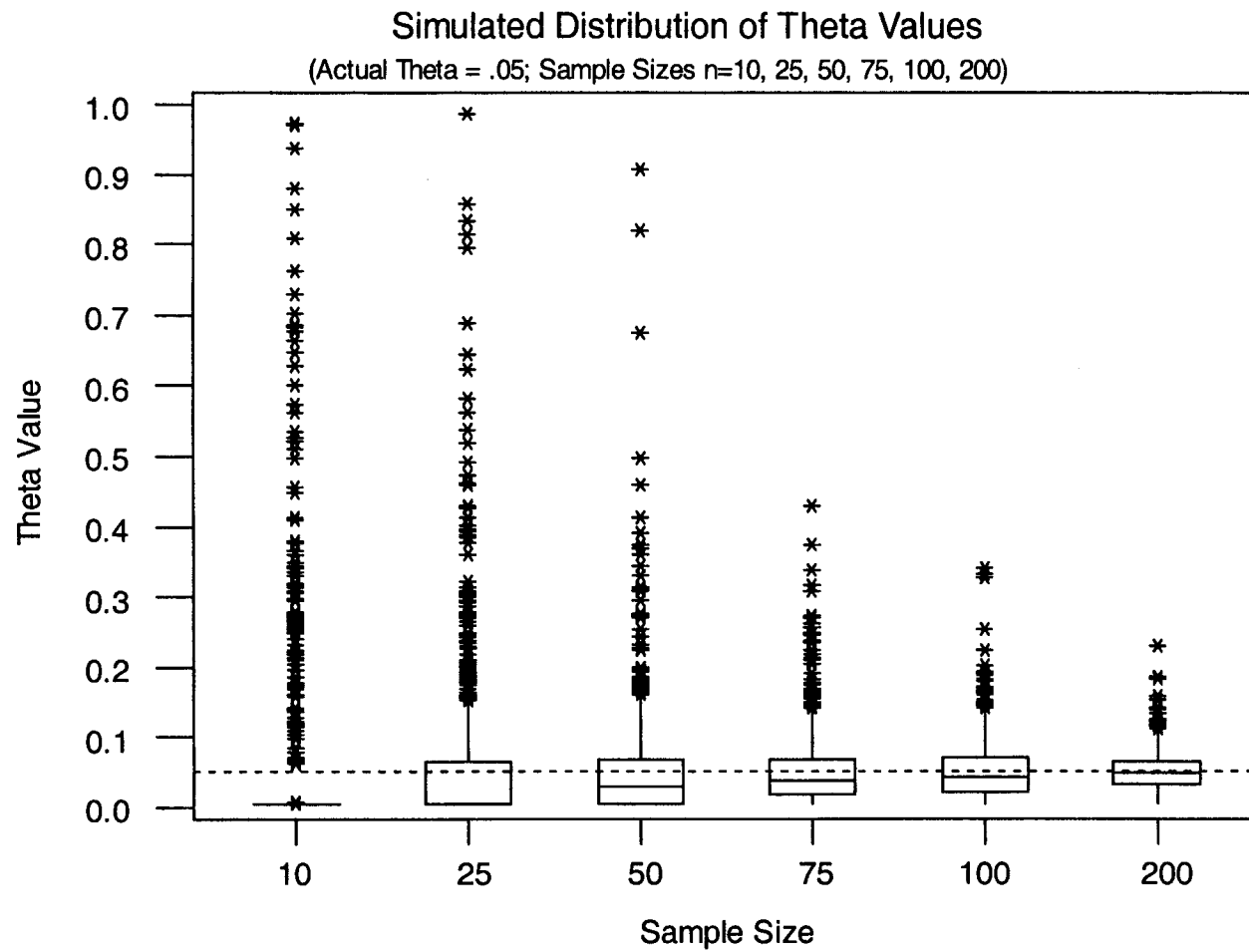


Figure 6. Simulated Distributions of $\hat{\theta}$ for $\theta = 1.0$, with Sample Sizes $n = 10, 25, 50, 75, 100$, and 200 .

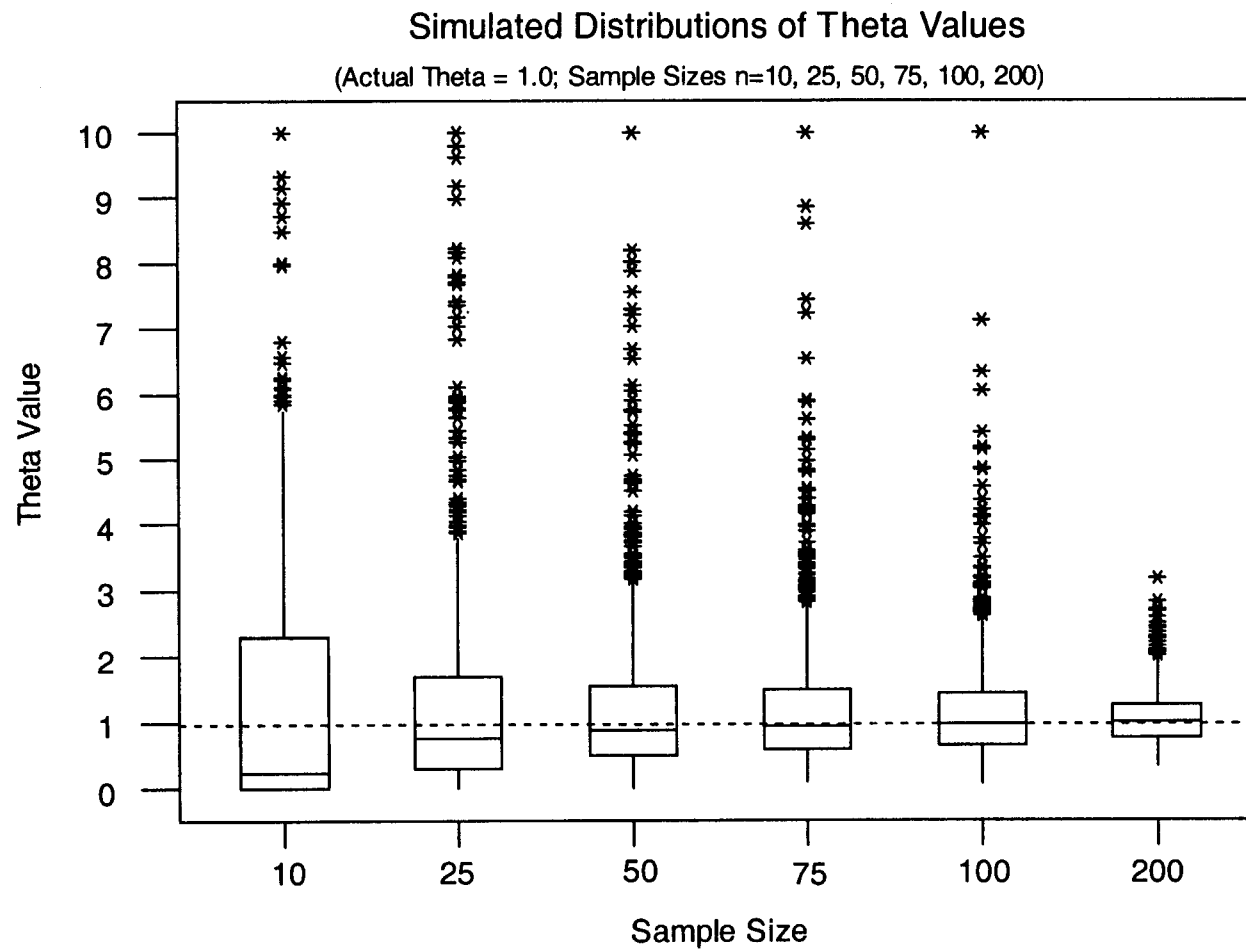


Figure 7. Simulated Distributions of $\hat{\theta}$ for $\theta = 5.0$, with Sample Sizes $n = 10, 25, 50, 75, 100$, and 200 .

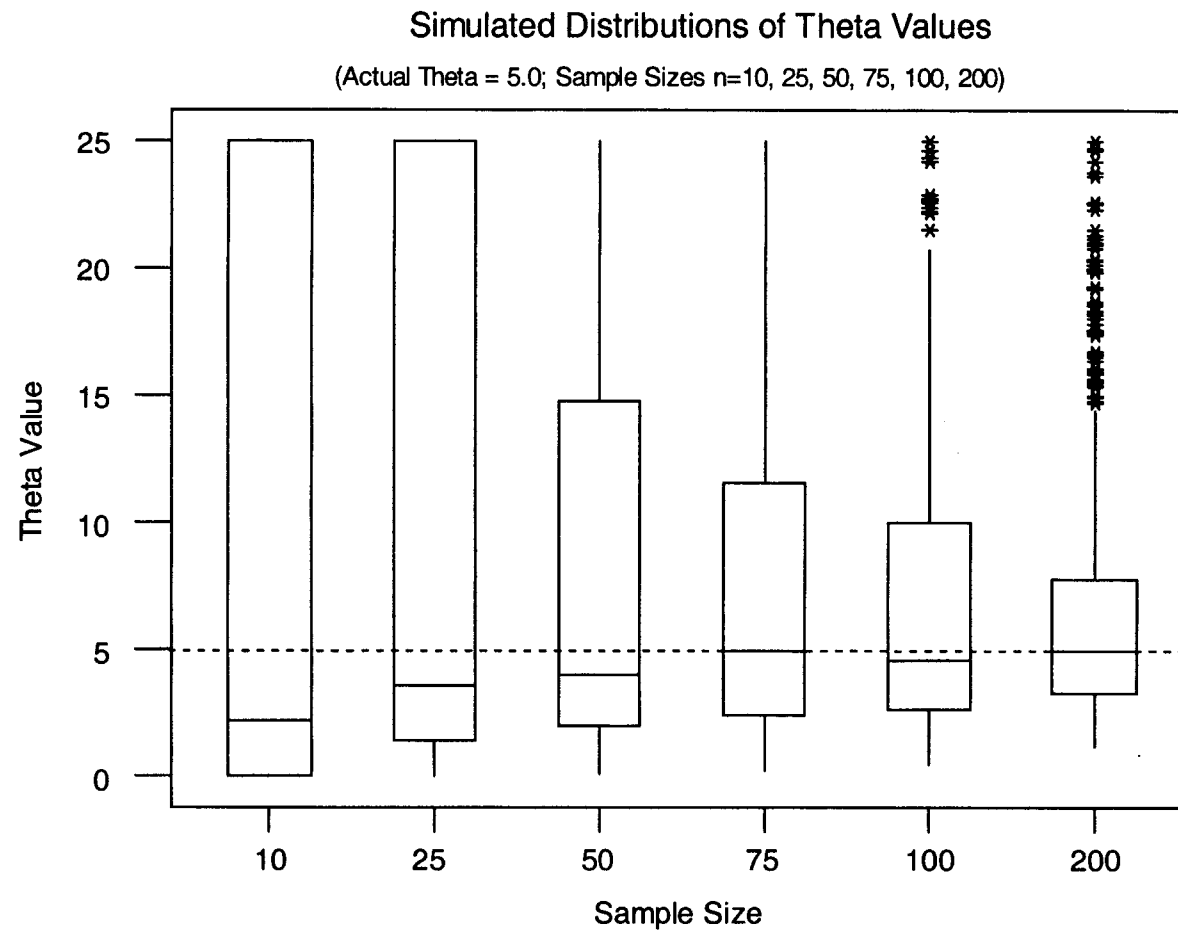


Figure 8. Battery Cathode Weights from a Low Volume Production Run.

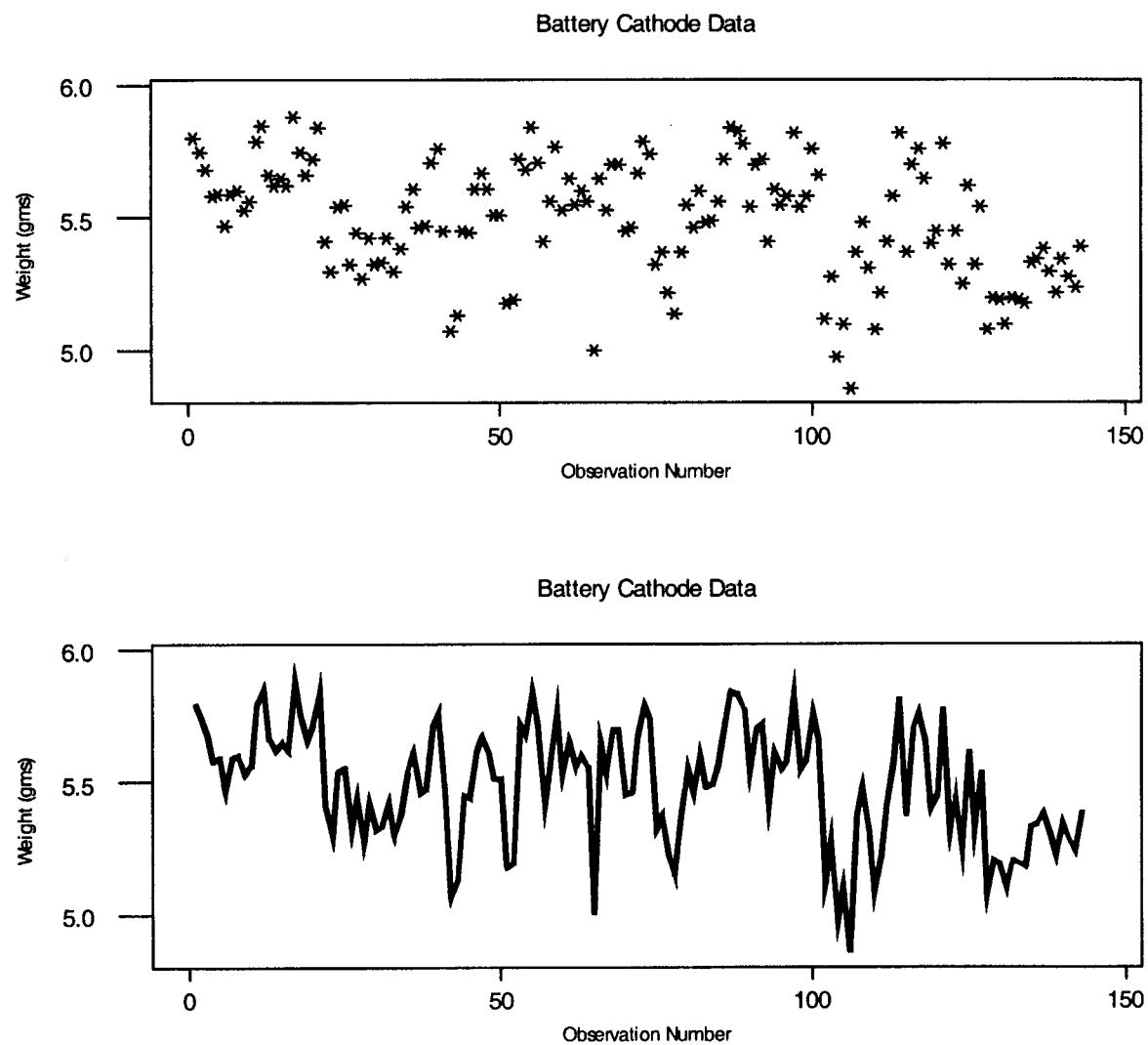


Figure 9. Battery Cathode Weights with Confidence Intervals for the μ_i 's ($n = 25$ and $n = 50$ data values).

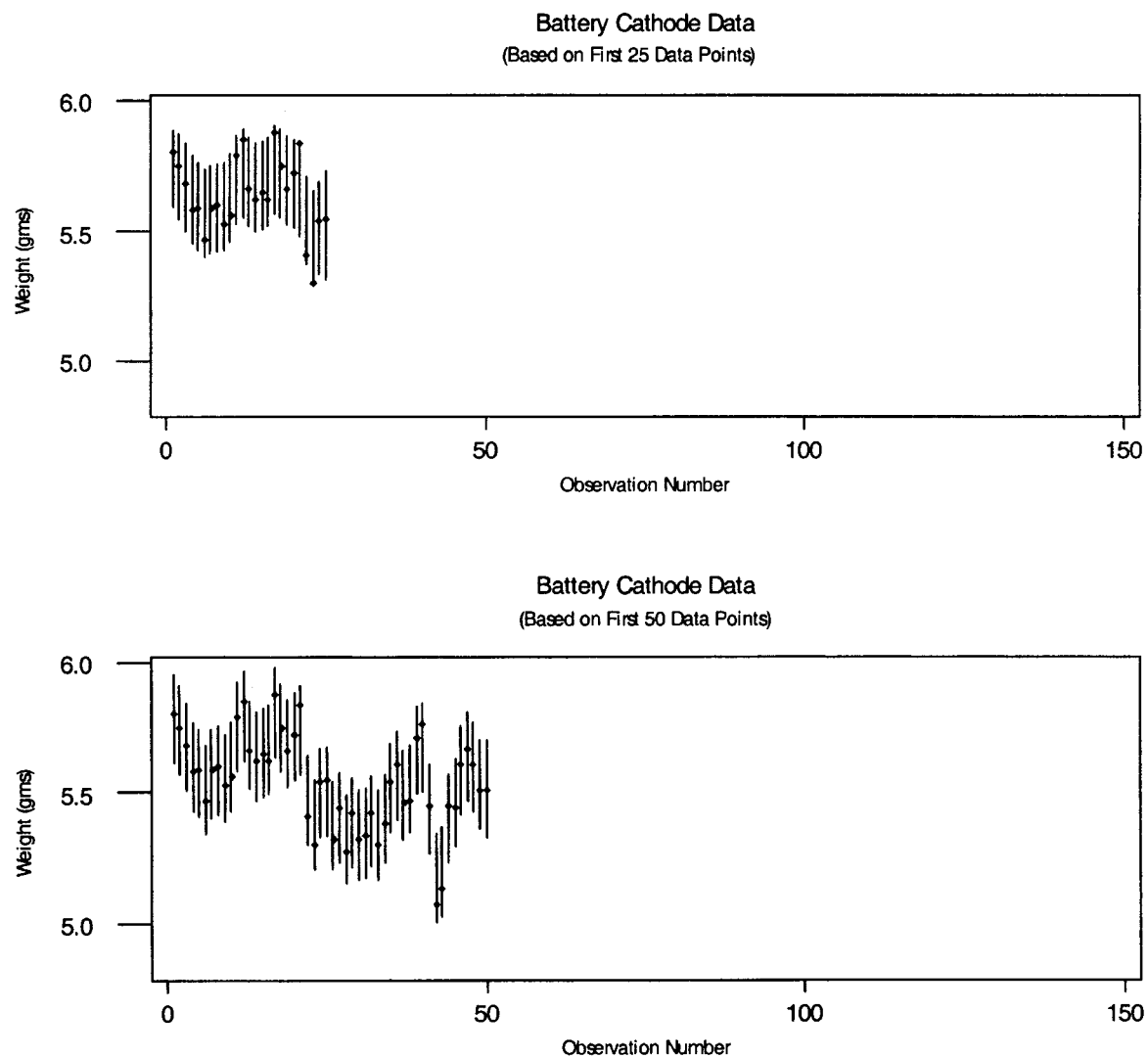


Figure 10. Battery Cathode Weights with Confidence Intervals for the μ_t 's ($n = 75$ and $n = 100$ data values).

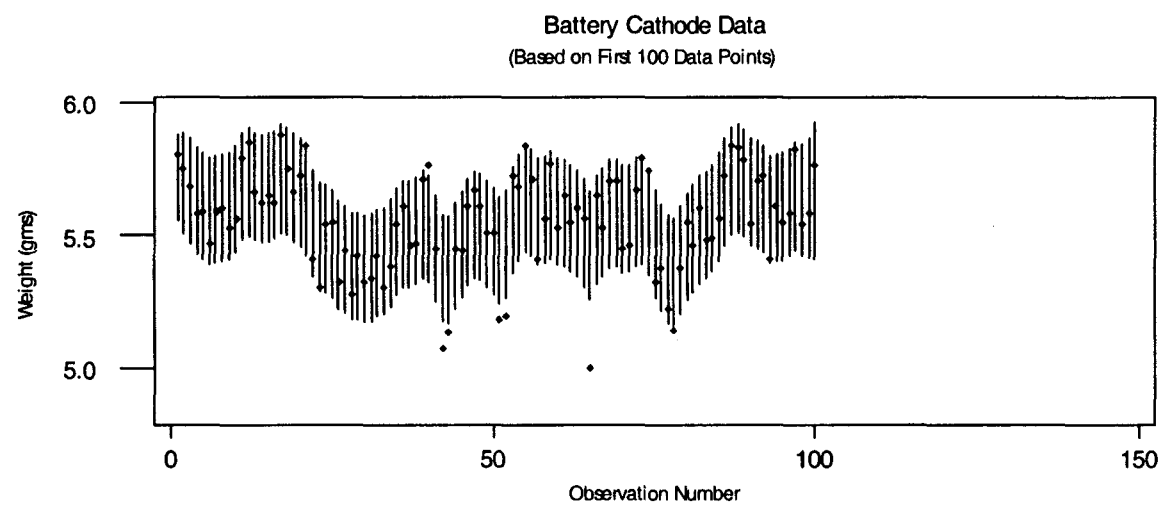
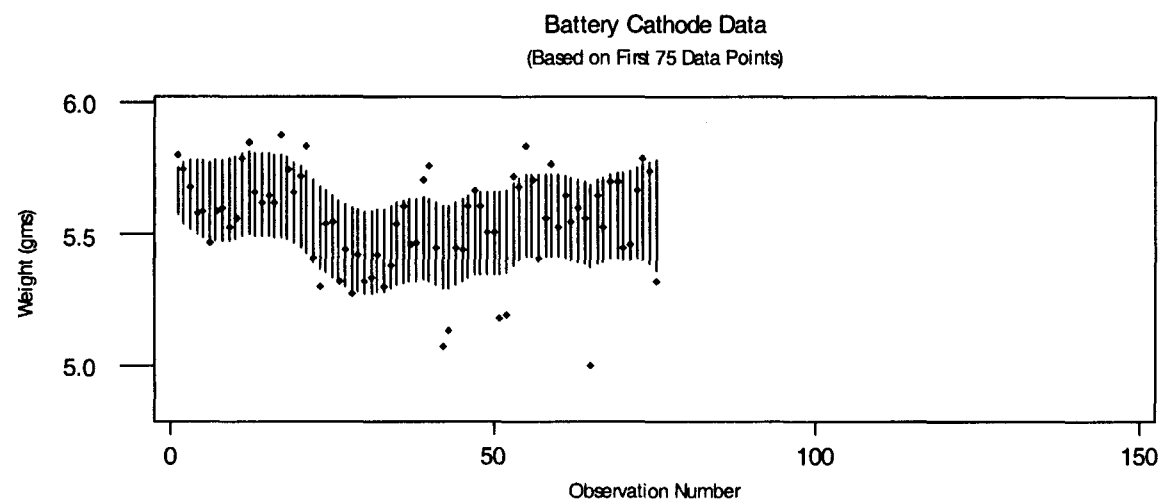


Figure 11. Battery Cathode Weights with Confidence Intervals for the μ_i 's ($n = 143$ data values).

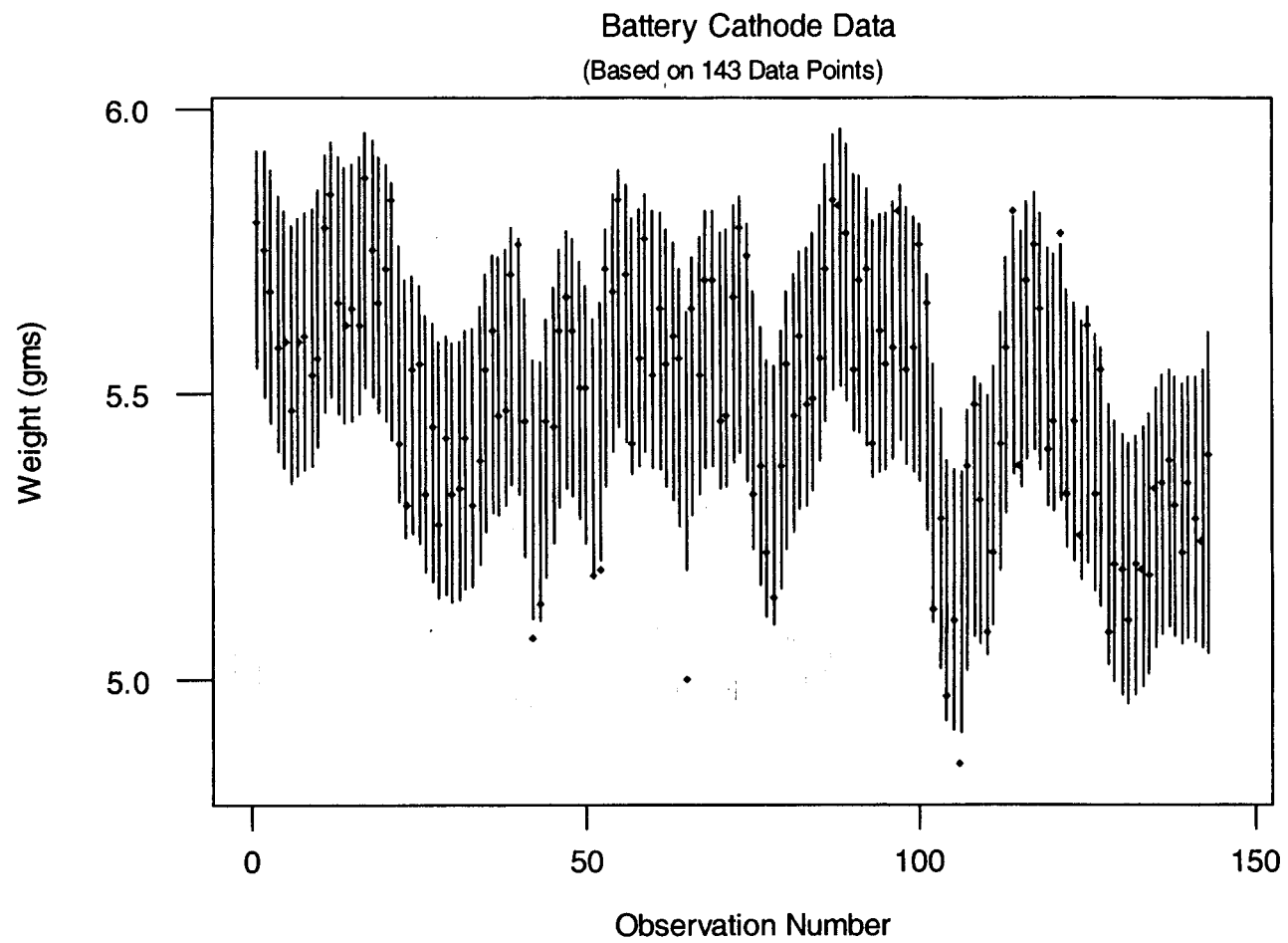
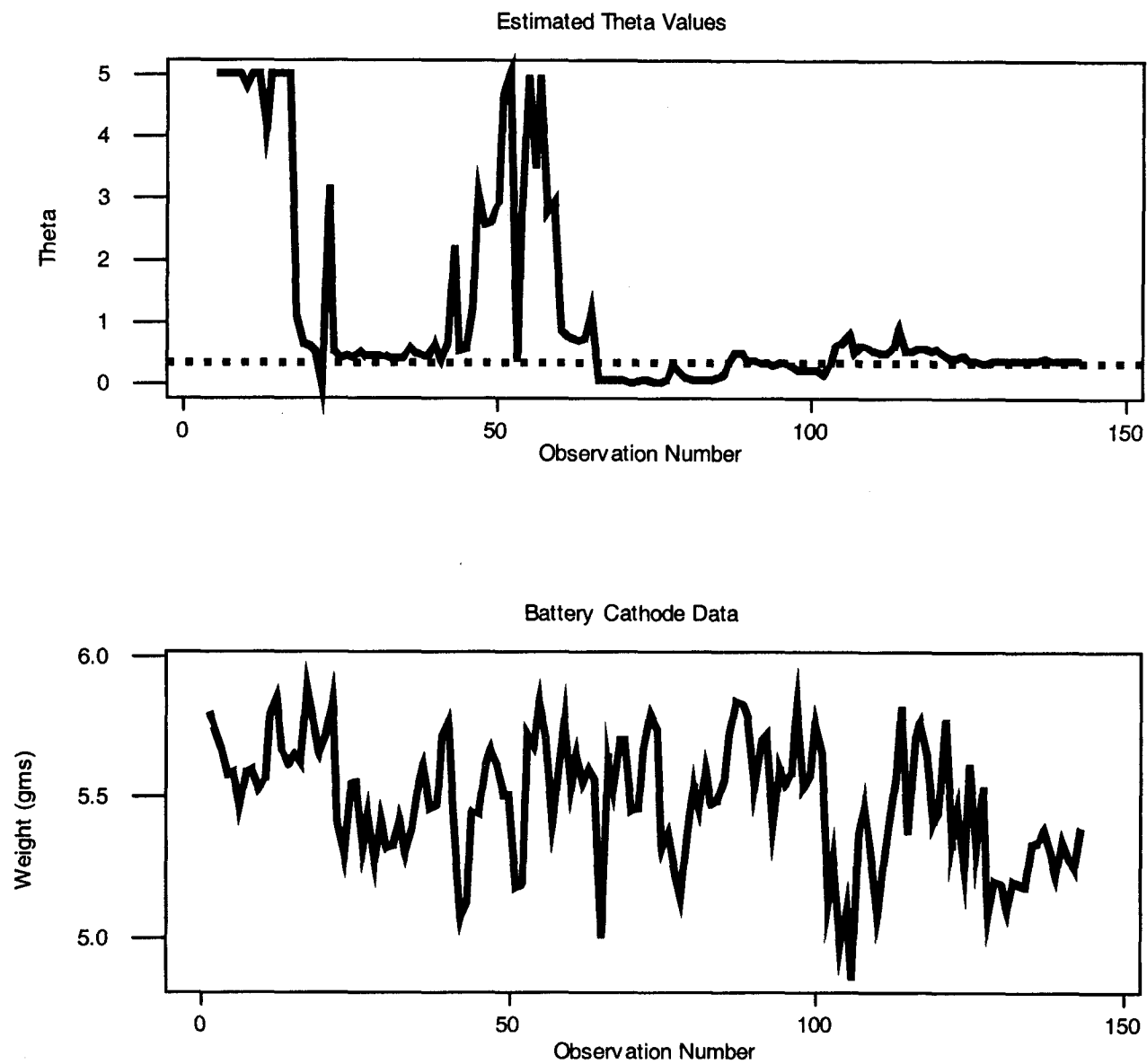


Figure 12. Estimates of θ for sample sizes $n = 3, 4, 5, \dots, 143$.



DISTRIBUTION:

1	MS0501	J. Spooner, 2338
1	MS0516	G. W. Smith, 1564
25	MS0829	S. V. Crowder, 12323
1	MS0829	K. V. Diegert, 12323
1	MS0829	K. S. Marlman, 12323
1	MS0830	E. W. Collins, 12335
1	MS0830	M. F. Martin, 12335
1	MS0957	R. C. Reuter, 1401
1	MS0960	J. Q. Searcy, 1400
1	MS9202	R. Bierbaum, 8116
1	MS9202	C. J. DeCarli, 8116 (Contractor)
1	MS9202	A. J. Johnson, 8116
1	MS9420	L. A. West, 8200
1	MS9018	Central Technical Files, 8940-2
2	MS0899	Technical Library, 4916
1	MS0612	Review & Approval Desk, 4912 For DOE/OSTI